DOCUMENT RESUME

ED 395 995 TM 025 194

| | |
|---|---|
| AUTHOR | Jiang, Hai; And Others |
| TITLE | An Estimation Procedure for the Structural Parameters of the Unified Cognitive/IRT Model. |
| PUB DATE | 8 Mar 96 |
| NOTE | 23p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New York, NY, April 9-11, 1996). |
| PUB TYPE | Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150) |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | *Cognitive Processes; *Estimation (Mathematics); *Item Response Theory; *Maximum Likelihood Statistics; Psychometrics; *Test Construction |
| IDENTIFIERS | *EM Algorithm |

ABSTRACT

L. V. DiBello, W. F. Stout, and L. A. Roussos (1993) have developed a new item response model, the Unified Model, which brings together the discrete, deterministic aspects of cognition favored by cognitive scientists, and the continuous, stochastic aspects of test response behavior that underlie item response theory (IRT). The Unified Model blends psychometric and cognitive science viewpoints and promises to allow the practitioner to recover cognitive information from simple, well-designed tests. This paper proposes an estimation procedure for the structural model parameters of the Unified Model that uses the marginal maximum likelihood estimation approach of Bock and Aitkin (1981) and the EM algorithm of A. P. Dempster, N. M. Laird, and D. B. Rubin (1977). In the maximization (M) step of the EM algorithm, because of the difficulties in computing the second derivative (Hessian) matrix and the possibility of multiple local maxima, using an alternative maximization procedure is proposed. This procedure, called Evolution Programming (Z. Michalewicz, 1994), has good properties in finding a global extremum. A simulation study is then given to show the effectiveness of the estimation procedure. (Contains 2 figures, 7 tables, and 16 references.) (Author/SLD)

ED 395 995

# An Estimation Procedure for the Structural Parameters of the Unified Cognitive/IRT Model

**Hai Jiang**
University of Illinois at Urbana-Champaign

**Louis DiBello**
Law School Admission Council

**William Stout**
University of Illinois at Urbana-Champaign

March 5, 1996



Paper presented at the 1996 NCME annual meeting, New York, NY, April 8, 1996.

## Abstract

DiBello, Stout, and Roussos (1993) have developed a new item response model, called the Unified Model, which brings together the discrete, deterministic aspects of cognition favored by cognitive scientists, and the continuous, stochastic aspects of test response behavior that underlie item response theory. The Unified Model blends psychometric and cognitive science viewpoints and promises to allow the practitioner to recover cognitive information from simple, well-designed tests.

In this paper, we propose an estimation procedure for the structural model parameters of the Unified Model that uses the marginal maximum likelihood estimation approach by Bock and Aitkin (1981) and utilizes the EM algorithm by Dempster, Laird, and Rubin (1977). In the maximization (M) Step of the EM algorithm, because of the difficulties in computing the second derivative (Hessian) matrix and the possibility of multiple local maxima, we propose using an alternative maximization procedure, called Evolution Programming (Michalewicz 1994), which has good properties in finding a global extremum. A simulation study is then given to show the effectiveness of our estimation procedure.



**Key words:** Unified Model, cognitive, item response theory, marginal maximum likelihood estimation, EM algorithm, Evolution Programming.

# 1 Introduction

DiBello, Stout, and Roussos (1993) have proposed a new psychometric approach to cognitive diagnostic assessment. They develop a new item response model, called the Unified Model, which brings together the discrete, deterministic aspects of cognition favored by cognitive scientists and the continuous, stochastic aspects of test response behavior that underlie item response theory. The Unified Model blends psychometric and cognitive science viewpoints and promises to allow the practitioner to recover cognitive information from simple, well-designed tests.

The ultimate goal of developing the Unified Model is to enable practitioners to cognitively classify the test takers and estimate their cognitive abilities, thereby extracting useful information about the test takers underlying cognitive processes on the test and their cognitive strengthes and weaknesses. To achieve this goal, it is essential to be able to estimate the model parameters before we can go on to classify examinees and estimate their abilities. Unfortunately, due to its structural complexity, until now there has been no estimation package available for the Unified Model. There are difficulties associated with the estimation problem, the main and foremost problem being the identifiability and estimability issues involving the model parameters. In this paper, we first give a brief overview of the Unified Model from the cognitive diagnosis viewpoint. After the overview, we will discuss briefly the relationship between the deterministically predicted ideal response patterns and the attribute states, as well as the identifiability and estimability issues involving the item parameters and the latent ability distribution parameters.

The third section of this paper concerns the estimation of the structural parameters of the Unified Model. In this section we propose an estimation procedure using the marginal maximum likelihood estimation approach by Bock and Aitkin (1981). Since we cannot maximize the marginal likelihood directly, we utilize the EM algorithm by Dempster, Laird, and Rubin (1977). In the maximization (M) step of the EM algorithm, because of the difficulties in computing the second derivative (Hessian) matrix and the possibility of multiple local maxima, we propose using Evolution Programming (Michalewicz, 1994), which has good properties in finding a global extremum.

In the fourth section of this paper, a simulation study is present showing the effectiveness of our proposed estimation procedure. We end the paper with a summary section.

# 2 The Unified Cognitive/IRT Model

## 2.1 Review of the Unified Model

The traditional role of tests in Education and Psychology is to rank the examinees and/or judge their proficiencies within a broad area of knowledge. For example, the GRE tests examinee proficiencies on verbal, quantitative, and analytical reasoning skills. In cognitive diagnosis, when giving a test, the test developers and administrators are interested not only in judging examinee proficiencies in a specific area of knowledge, but also in getting information on examinees' underlying cognitive processes used on the test. This happens in the usual classroom setting: when giving a test, a teacher not only wants to know which grade Johnny gets, but perhaps more importantly, she wants to know whether Johnny has really mastered the Algebraic Rules of Exponents. In other words, she wants to assess the examinee mastery on a variety of cognitive attributes. An attribute represents a cognitive quality required for solution of a test item: it can be anything based on the procedures, skills, processes, strategies, or the knowledge that an examinee needs to possess to solve the item.

There are two distinct approaches in cognitive diagnosis: the continuous multidimensional latent trait approach favored by psychometricians and the discrete approach favored by cognitive scientists. In the usual latent trait approach, a few broadly described continuous latent traits are postulated to account for systematic examinee response behavior on a test. As Snow and Lohman (1989) noted, this approach has a weak cognitive foundation. Although it sometimes sounds like the multidimensional underlying latent traits are cognitive in nature, it is generally agreed that this approach has only been successful with broad, composite abilities. In particular, for example, it is of little help in trying to determine specific cognitive characteristics of examinees for the purpose of instruction.

For the discrete approach, an example is latent class analysis (see for example, Lazersfeld and Henry, 1968). A latent class analysis involves the postulation of a number of latent classes. In a latent class analysis, examinee ability is *not* represented as a continuous variable on dimensions defined by the cognitive components. Instead, it is modeled by a vector of 1s and 0s indicating for each cognitive component whether an examinee does or does not possess the skills needed for successful performance on the component. Latent class analyses either involve a large number of classes so that it is infeasible for estimation, or there are only a few latent classes that the results are similar to multidimensional latent trait analyses in their coarseness of latent structure assumed (Bock and Aitkin, 1981; Bartholomew, 1987; Takane and de Leeuw, 1987; Haertel, 1990).

The Unified Model approach blends psychometric and cognitive science viewpoints. It is based upon a new item response model, called the Unified Model. Below we will give a brief review of the Unified Model.

Following Tatsuoka (Tatsuoka, K.K., 1984, 1985, 1990; Tatsuoka, K.K. and Tatsuoka, M.M., 1987), we consider a test of length $I$ with $K$ postulated cognitive attributes, and a matrix $\mathbf{Q} = (q_{ki})_{K\times I}$, where

$$q_{ki} = \begin{cases} 1 & \text{if item } i \text{ requires attribute } k \\ 0 & \text{if not} \end{cases}$$

The $K$ attributes include those of interest for cognitive diagnosis, as well as others inadvertently present in the test. The $\mathbf{Q}$ matrix specifies which attributes must be mastered in order to correctly answer each item.

The $\mathbf{Q}$ matrix represents a presumed choice of strategy for each item. By strategy we mean the steps that are used in answering the item.

Let $\underline{\alpha} = (\alpha_1, ..., \alpha_K)^T$ be a vector denoting an examinee's attribute state, where $\underline{x}^T$ denotes the transpose of vector $\underline{x}$, and

$$\alpha_k = \begin{cases} 1 & \text{if examinee has mastered attribute } k \\ 0 & \text{if not} \end{cases}$$

A given examinee attribute state $\underline{\alpha} = (\alpha_1, ..., \alpha_K)^T$, along with the $\mathbf{Q}$ matrix produces *the ideal response pattern associated with* $\underline{\alpha}$ *and determined by* $\mathbf{Q}$:

$$x_{\mathbf{Q}}(\underline{\alpha}) = (x_1, ..., x_I)^T \tag{1}$$

It is defined as follows: for the ideal response, item $i$ is answered correctly if the examinee possesses all the attributes as required by $\mathbf{Q}$ for this item; otherwise, item $i$ is answered incorrectly. In mathematical terms,

$$x_i = \begin{cases} 0 & \text{if there is an attribute } k \text{ for which } q_{ki} = 1 \text{ but } \alpha_k = 0 \\ 1 & \text{if not} \end{cases} \tag{2}$$

In reality, examinee responses are seldom consistent with such a simple deterministic model. We expect examinee responses to differ from the ideal response patterns. The Unified Model approach models the probabilistic variation in examinee responses by incorporating the following four major sources of response variation.

Strategy: The examinee mav use a different strategy from that presumed by the Q matrix.

Completeness: An item may require attributes that are not listed in the **Q** matrix. If so, we will say the **Q** matrix is incomplete for the it.m.

Positivity: In some cases, an examinee who possesses an attribute will fail to apply it correctly to an item, and another examinee who lacks the attribute will apply it correctly to the item. If such response behavior is prevalent among the examinees, we will say the attribute is low positive for the item.

Slips: The examinee may commit a random error.

To allow for cases in which multiple strategies are used by examinees and cases in which the **Q** matrix is incomplete, the notion of a latent residual ability $\eta$ is in. duced in the Unified Model. Hence under the Unified Model, the complete latent ability for an examinee is $\underline{\theta} = (\eta, \underline{\alpha}^T)^T = (\eta, \alpha_1, ..., \alpha_K)^T$, where $\underline{\alpha}$ is the examinee's attribute state and $\eta$ is his residual ability.

Under the Unified Model, the probability of an examinee answering item $i$ correctly (denoted $Y_i = 1$) given that he has ability $(\eta, \underline{\alpha}^T)^T$ is

$$P_i(\eta, \underline{\alpha}; \underline{\beta}_i) = P(Y_i = 1|\eta, \underline{\alpha}; \underline{\beta}_i) = (1-p)[d_i S_{\underline{\alpha},i} P_{b_i}(\eta + 2c_i) + (1-d_i) P_{b_i}(\eta)] \tag{3}$$

where

$p$ = probability of a random slip

$d_i$ = probability of selecting **Q** strategy for item $i$

$c_i$ = completeness index of **Q** matrix for item $i$

$$S_{\underline{\alpha},i} = \prod_{k=1}^{K} (\pi_{ki})^{q_{ki}\alpha_k}(r_{ki})^{q_{ki}(1-\alpha_k)} = P(\text{applying required attributes correctly to } i|\underline{\alpha})$$

$$\pi_{ki} = P(\text{applying attribute } k \text{ correctly to } i|\alpha_k = 1)$$

$$r_{ki} = P(\text{applying attribute } k \text{ correctly to } i|\alpha_k = 0)$$

$$P_{b_i}(\eta) = \frac{1}{1+e^{-1.701(\eta-b_i)}} = 1 \text{ parameter logistic (1PL) with difficulty } b_i$$

$$\underline{\beta}_i = (b_i, c_i, d_i, \pi_{1i}, \ldots \pi_{Ki}, r_{1i}, \ldots, r_{Ki})^T, \qquad i = 1, \ldots, I$$

The four sources of response variation are incorporated in the Unified Model through the parameters $p$, $c_i$'s, $d_i$'s, the $\pi$'s and the $r$'s, for example, the $\pi$'s and $r$'s are used to model the positivity of the attributes. For a derivation of the model, the reader is directed to DiBello, Stout, and Roussos (1993).

From now on, we assume $p = 0$ (no slip) for all items. Thus the probability of answering item $i$ correctly given examinee ability $(\eta, \underline{\alpha}^T)^T$ becomes

$$P_i(\eta, \underline{\alpha}; \underline{\beta}_i) = d_i S_{\underline{\alpha},i} P_{b_i}(\eta + 2c_i) + (1 - d_i) P_{b_i}(\eta) \qquad (4)$$

## 2.2 Ideal response patterns

A given examinee attribute state $\underline{\alpha}$, along with the **Q** matrix produces the ideal response pattern associated with $\underline{\alpha}$ as defined by (1) and (2). Since we postulate $K$ attributes, there are $2^K$ different attribute states. The number of different ideal response patterns, however, is usually smaller than $2^K$ because of the fact that different attribute states may produce exactly the same ideal response pattern.

*Example 1:* For the **Q** matrix given below,

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

both attribute states $(1, 0, 1, 1, 0)^T$ and $(1, 0, 0, 1, 0)^T$ produce the same ideal response pattern $(0, 0, 0, 0, 0, 0, 1, 1, 0, 0)^T$.

*Definition 1:* For two attribute states $\underline{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_{1K})^T$ and $\underline{\alpha}_2 = (\alpha_{21}, \ldots, \alpha_{2K})^T$, $\underline{\alpha}_1$ is a substate of $\underline{\alpha}_2$ and is denoted by $\underline{\alpha}_1 \leq \underline{\alpha}_2$, if $\alpha_{1k} \leq \alpha_{2k}$ for $k = 1, \ldots, K$.

In example 1, attribute state $(1,0,0,1,0)^T$ is a substate of attribute state $(1,0,1,1,0)^T$.

*Definition 2:* Among all the attribute states that produce the same ideal response pattern, the canonical state is the one that has the smallest number of 1's.

In example 1, $(1,0,0,1,0)^T$ is the canonical state that produces the ideal response pattern $(0,0,0,0,0,0,1,1,0,0)^T$.

It can be shown that the canonical state as given by the above definition is unique.

*Definition 3:* An attribute state $\underline{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ is a direct sum of two attribute states $\underline{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_{1K})^T$ and $\underline{\alpha}_2 = (\alpha_{21}, \ldots, \alpha_{2K})^T$, denoted by $\underline{\alpha} = \underline{\alpha}_1 \vee \underline{\alpha}_2$, if

$$\alpha_k = \alpha_{1k} \vee \alpha_{2k} = max(\alpha_{1k}, \alpha_{2k}), \qquad \text{for } k = 1, \ldots, K$$

Whether an attribute state is a canonical state can be determined by the following proposition.

*Proposition 1:* An attribute state is a canonical state, if

- it is the attribute state of all 1's or the attribute state of all 0's,
- it is a column of the **Q** matrix, or
- it has substates that are columns of the **Q** matrix and it is the direct sum of these substates.

*Example 2:* For the **Q** matrix given below,

$$\mathbf{Q} = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

attribute state $(0,1,1,1)^T$ is a canonical state, because it has substates $(0,1,1,0)^T$ and $(0,0,1,1)^T$ that are columns of **Q**, and the direct sum of these two substates is $(0,1,1,1)^T$.

As results of the above proposition, we have the following corollaries concerning the number of canonical states.

*Corollary 1:* If among the $K$ attributes postulated, only $K'$ are required by all the items, there will be at most $2^{K'}$ canonical states.

In this case, there are $K - K'$ attributes not required by any item; in other words, they are redundant. Below we assume this situation never happens.

*Corollary 2:* Consider all the items each requiring a single attribute, if the number of different attributes required by these items is $K'$, there will be at least $2^{K'}$ canonical states.

The canonical states can now be used as representatives of attribute classes, so we can index the set of all ideal response patterns, or the set of attribute classes by $l = 1, \ldots, L$, and replace attribute state $\underline{\alpha}$ by index $l$ in our notation heretofore. The latent space can now be thought of as $\{(\eta, l)^T : \eta \in \mathbf{R}, l = 1, ..., L\}$.

For the distribution of latent ability $(\eta, l)^T$, we assume in our model a finite mixture of normals with the mixing probability $p_l$ and $N(\mu_l, \sigma^2)$ for given $l$, i.e., the density of latent ability distribution

$$\pi(\underline{\theta}; \underline{\phi}) = \pi(\eta, l; \underline{\phi}) = p_l \frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{(\eta - \mu_l)^2}{2\sigma^2}] \tag{5}$$

where the latent ability distribution parameters

$$\underline{\phi} = (p_1, \ldots, p_L, \mu_1, \ldots, \mu_L, \sigma^2)^T$$

## 2.3 Log likelihood function

Suppose there are a total of $N$ examinees. Let $\mathbf{B} = (\underline{\beta}_1, \ldots, \underline{\beta}_I)$ be the totality of item parameters, and assume the latent space is complete with respect to the latent ability vector $\underline{\theta}$ so that the local independence given $\underline{\theta}$ holds

$$P(\underline{Y}_n|\underline{\theta}; \mathbf{B}) = \prod_{i=1}^{I} P(Y_{ni}|\underline{\theta}; \underline{\beta}_i) = \prod_{i=1}^{I} P_i(\underline{\theta}; \underline{\beta}_i)^{Y_{ni}}(1 - P_i(\underline{\theta}; \underline{\beta}_i))^{1-Y_{ni}}$$

where $\underline{Y}_n$ is the response vector for examinee $n$.

The marginal likelihood function, which is the likelihood function given the response matrix $\mathbf{Y}$ and the matrix of latent abilities $\mathbf{\Theta}$ integrated over the latent ability distribution, is given by

$$L(\mathbf{B}, \underline{\phi}|\mathbf{Y}) = \int L(\mathbf{B}, \underline{\phi}|\mathbf{Y}, \mathbf{\Theta}) F(d\mathbf{\Theta}) = \prod_{n=1}^{N} \int P(\underline{Y}_n|\underline{\theta}; \mathbf{B}) \pi(\underline{\theta}; \underline{\phi}) d\underline{\theta}$$

Here we have used the independence of examinees to factor the likelihood function $L(\mathbf{B},\underline{\phi}|\mathbf{Y},\Theta)$ and to factor $F(d\Theta)$ into a product measure.

Taking the logarithm and using (5), the log likelihood function given $\mathbf{Y}$ is

$$\ln L(\mathbf{B},\underline{\phi}|\mathbf{Y}) = \sum_{n=1}^{N} \ln \sum_{l=1}^{L} \int P(\underline{Y}_n|\eta,l;\mathbf{B})p_l \frac{1}{\sqrt{2\pi}\sigma} exp[-\frac{(\eta-\mu_l)^2}{2\sigma^2}]d\eta \qquad (6)$$

## 2.4 Some model identifiability and estimability issues

There are various identifiability and estimability problems involving the parameters of the Unified Model. First we give two definitions, which deal with two different causes for being unable to estimate model parameters.

*Definition 4:* If a probability model $p(\underline{y}|\underline{\phi})$ is parametrized by a vector $\underline{\phi}$, we say the model parameter $\underline{\phi}$ is not identifiable if there exists a $\underline{\phi}' \neq \underline{\phi}$ such that for all $\underline{y}$

$$p(\underline{y}|\underline{\phi}) = p(\underline{y}|\underline{\phi}')$$

i.e., the distribution of $\underline{y}$ is the same for $\underline{\phi}$ and $\underline{\phi}'$. In this case, the data simply cannot yield any information to distinguish $\underline{\phi}$ from $\underline{\phi}'$. Further, if $m$ constraints (e.g., fixing $m$ components of $\underline{\phi}$) render the $n$-dimensional $\underline{\phi}$ identifiable, we say that $n-m$ of the components are identifiable.

*Definition 5:* If a model $p(\underline{y}|\underline{\phi})$ is parametrized by a vector $\underline{\phi} = (\phi_1,\ldots,\phi_n)^T$, we say a component $\phi_k$ of the model parameter $\underline{\phi}$ is not estimable if the model does not actually involve the component. In this case, the data simply cannot yield any information about $\phi_k$.

First, let us consider an identifiability problem that arises involving the $\pi$'s and $r$'s.

*Proposition 2:* For item $i$, let $K_i = \sum_{k=1}^{K} q_{ki}$, the number of attributes required by $i$. Then among the $2K_i$ of $\pi$'s and $r$'s for which $q_{ki} = 1$, only $K_i + 1$ of them are identifiable.

This identifiability problem is caused by the nonlinear constraints among the $S_{\underline{\alpha},i}$'s resulting from their being products of the $\pi$'s and $r$'s. For illustrative purposes, suppose there are just two attributes and item 1 requires both. Recall (3) and consider its $S$'s. Then we have four $\pi$'s and $r$'s ($\pi_{11},\pi_{21},r_{11},r_{21}$) to be estimated, or equivalently four $S$'s to be estimated ($S_{11} = \pi_{11}\pi_{21}$, $S_{21} = \pi_{11}r_{21}$, $S_{31} = \pi_{21}r_{11}$, and $S_{41} = r_{11}r_{21}$. Here the first index of the $S$'s denotes the attribute states $(1,1)^T$, $(1,0)^T$, $(0,1)^T$, and $(0,0)^T$, respectively). Since $S_{41} = S_{21}S_{31}/S_{11}$, only three $S$'s are identifiable, or equivalently only three of the four $\pi$'s and $r$'s are identifiable.

To resolve the above identifiability problem involving the $\pi$'s and $r$'s, if an item $i$ requires $K_i$ attributes, we will fix the first $K_i - 1$ $\pi$'s at 1, leaving only the last $\pi$ free, so that item $i$ now has only $K_i + 1$ free $\pi$ and $r$ parameters.

Next, let us look at the identifiability issue involving the $b_i$ and $\mu_l$, recalling (3) and (5).

*Proposition 3:* If holding all other parameters fixed, and adding the same constant to every $b_i$ and every $\mu_l$, the log likelihood function $\ln L(\mathbf{B}, \underline{\phi}|\mathbf{Y})$ will not change. Note that this is of course the usual identifiability problem occurring with ordinary IRT logistic modeling.

Proofs of the above results can be found in Jiang (1996). There are other identifiability and estimability problems; below we give some examples. If $d_i = 1$ (i.e., we are certain the $\mathbf{Q}$ strategy for item $i$ will be selected by all examinees), the Unified Model for the $i$th item becomes

$$P_i(\eta, \underline{\alpha}; \underline{\beta}_i) = S_{\underline{\alpha}, i} P_{b_i}(\eta + 2c_i)$$

Since .

$$P_{b_i}(\eta + 2c_i) = P_{b_i - 2c_i}(\eta)$$

we cannot estimate $b_i$ and $c_i$ separately when $d_i = 1$. In the sense of Definition 4, $b_i$ and $c_i$ are unidentifiable when $d_i = 1$, because different sets of $(b_i, c_i)$ with the same value at $b_i - 2c_i$ will give the same $P_i(\eta, \underline{\alpha}; \underline{\beta}_i)$. Similarly, we can argue that if $d_i$ is close to 1, we will not have enough information from the data to accurately estimate $b_i$ and $c_i$ separately, but rather can estimate them together through the linear combination $b_i - 2c_i$.

If $d_i = 0$, then

$$P_i(\eta, \underline{\alpha}; \underline{\beta}_i) = P_{b_i}(\eta)$$

In this case, we cannot estimate $S_{\underline{\alpha} i}$ for any possible $\underline{\alpha}$, nor can we estimate $c_i$ (i.e., $S_{\underline{\alpha} i}$ and $c_i$ are not estimable in the sense of Definition 5).

Since $d_i$ is the probability of selecting $\mathbf{Q}$ strategy for item $i$, we can normally assume $d_i$ is bounded away from 0, unless the $\mathbf{Q}$ matrix under consideration is badly constructed from cognitive perspective.

# 3 Estimating the Structural Parameters of the Unified Model

Since directly maximizing $\ln L(\mathbf{B}, \underline{\omega}|\mathbf{Y})$ over $\mathbf{B}$ and $\underline{\omega}$ is infeasible, we use the EM algorithm.

## 3.1 EM algorithm for the Unified Model

The EM algorithm, as its name suggests, is divided into two steps: the E (expectation) step, and the M (maximization) step. Cyclical application of the E step and the M step continues till a certain convergence criterion is met.

In the E step, the conditional expectation of log likelihood of complete data given the incomplete data and current parameter estimates is computed. In our study, the incomplete data is the observed response matrix $\mathbf{Y}$ and the complete data is the responses plus the examinee latent ability matrix $\mathbf{\Theta}$. So in the E step, the following quantity is computed

$$Q(\mathbf{B}, \underline{\omega}; \mathbf{B}', \underline{\omega}') = E[\ln L(\mathbf{B}, \underline{\omega}|\mathbf{Y}, \mathbf{\Theta})|\mathbf{Y}; \mathbf{B}', \underline{\omega}']$$

where the expectation is taken with respect to $\mathbf{\Theta}$. Here $\mathbf{B}'$ and $\underline{\omega}'$ are the parameter estimates resulted from the M step in the previous iteration. Here and below we follow the standard notation in the literature of EM algorithm. It is understood that the $Q$ functions in the E and M steps depend on the observed response matrix $\mathbf{Y}$.

It turns out that for the Unified Model the following decomposition holds

$$Q(\mathbf{B}, \underline{\omega}; \mathbf{B}', \underline{\omega}') = Q_0(\underline{\omega}; \mathbf{B}', \underline{\omega}') + \sum_{i=1}^{I} Q_i(\underline{\beta}_i; \mathbf{B}', \underline{\omega}')$$

where $Q_0(\underline{\omega}; \mathbf{B}', \underline{\omega}')$ involves only the ability distribution parameter $\underline{\omega}$ and each $Q_i(\underline{\beta}_i; \mathbf{B}', \underline{\omega}')$ involves only the item structure parameter $\underline{\beta}_i$ for item $i$.

In the M step, $Q(\mathbf{B}, \underline{\omega}; \mathbf{B}', \underline{\omega}')$ is maximized over the parameters $\mathbf{B}$ and $\underline{\omega}$ for given $\mathbf{B}'$, $\underline{\omega}'$ and $\mathbf{Y}$. Because of the decomposition in the E step, we can separately maximize $Q_0(\underline{\omega}; \mathbf{B}', \underline{\omega}')$ over $\underline{\omega}$ and each $Q_i(\underline{\beta}_i; \mathbf{B}', \underline{\omega}')$ over $\underline{\beta}_i$.

While there exists a closed-form solution $\hat{\underline{\omega}}$ for maximizing $Q_0(\underline{\omega}; \mathbf{B}', \underline{\omega}')$ over $\underline{\omega}$, no closed-form solution exists for maximizing $Q_i(\underline{\beta}_i; \mathbf{B}', \underline{\omega}')$ over $\underline{\beta}_i$. Because of the difficulties in computing the derivatives, especially the second derivative (Hessian) matrix, and the possibility of many local maxima, the conventional Newton-Raphson method is not feasible to apply. So instead of using Newton-Raphson, we use Evolution Programming (Michalewicz, 1994) to maximize each $Q_i(\underline{\beta}_i; \mathbf{B}', \underline{\phi}')$ over $\underline{\beta}_i$.

Any optimization task can be thought of as a search through a space of potential solutions. Evolution Programming is a stochastic algorithm whose search method emulates the natural phenomena of genetic inheritance and Darwinian strife for survival. An Evolution Programming maintains a population of individuals $P(t) = \{x_1^t, ..., x_n^t\}$ for use in iteration $t$. Each individual is a vector and represents a potential solution to the problem at hand (i.e., a potential optimizer of the problem). Each solution $x_i^t$ is evaluated to give some measure of " fitness". Then, as a result of iteration $t$ a new population $P(t+1)$ for use in iteration $t+1$ is formed by selecting the more "fit" individuals (the select step). Some members of this new population undergo transformations (alter step) by means of "genetic" operators to form new potential solutions. There are unary transformations (mutation type), which create new individuals by a small change in a single individual, and higher order transformations (crossover type), which create new individuals by combining segments from several (two or more) individuals. After several generations the program converges with the goal being that the best individual in this final generation represents a near-optimum solution.

For our problems at hand, since we are maximizing each $Q_i(\underline{\beta}_i; \mathbf{B}', \underline{\phi}')$ over $\underline{\beta}_i$, the population of individuals are vectors of possible values for $\underline{\beta}_i$. One iteration consists of operations such as mutation, crossover, and selection. Evolution continues through generations until a certain convergence accuracy is obtained. Then the maximizer $\hat{\underline{\beta}}_i$ of $Q_i(\underline{\beta}_i; \mathbf{B}', \underline{\phi}')$ is given by the best solution vector of the final generation.

# 4 Simulation Study and Results

We use a carefully constructed **Q** matrix with 5 postulated attributes, and an examinee population having a selected subset of all 32 possible attribute states to demonstrate the effectiveness of our estimation procedure.

We postulate 10 "core" items. The **Q** matrix and the item parameter settings are given in Table 1. The rationale of choosing the item parameter setting as given in Table 1 is that the completeness index $c_i$ of the **Q** matrix is moderate to high. That is, it is relatively a single strategy test (i.e., $d_i$'s are close to 1 so that we are fairly sure an examinee will choose the single strategy postulated by the **Q** matrix), and the positivity is high so that the $\pi$'s are close to 1 while the $r$'s are small. With a set of well defined attributes and reasonably well constructed **Q** matrix, our choice of the item parameter setting is quite plausible. The simulated test consists of 40 items obtained by replicating each core item 4 times. For the **Q** matrix postulated, there are a total of 24 possible attribute classes (i.e., there are only 24 different ideal response patterns resulting in 24 attribute classes from the 32 attribute states). 1000 examinees are generated by assuming that only 10 of the 24 attribute classes actually occur in the examinee population. Table 2 gives the latent ability distribution parameter settings for the 10 classes, along with their representative canonical states and their ideal response patterns. In Table 2 and subsequent tables, the column lable IR refers to ideal response pattern number. Note that while our choice of the $\mu_l$ is somewhat arbitrary, the $\mu_l$ ordering is consistent with the partial ordering existing among the attribute states. Because the empirical work needed to find "realistic" model parameter values ($\mathbf{B}$, $\underline{\rho}$) has not been done, we have been forced to select what seem to be plausible values for the model parameters.

Recall that the Unified Model for the $i$th item is given by (4). The estimated item and latent ability distribution parameters as a result of our EM algorithm run are given in Tables 3 and 4, respectively.

Table 1. **Q** matrix and true item parameters

| item | | attr. 1 | attr. 2 | attr. 3 | attr. 4 | attr. 5 | $b_i$ | $c_i$ | $d_i$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $\pi$ | 0.9 | 0.8 | | | | −0.3497 | 0.90 | 0.90 |
| | $r$ | 0.1 | 0.4 | | | | | | |
| 2 | $\pi$ | | | 0.8 | 1.0 | 0.9 | −0.1929 | 0.70 | 0.95 |
| | $r$ | | | 0.4 | 0.3 | 0.1 | | | |
| 3 | $\pi$ | | 1.0 | 0.8 | 0.8 | | −0.0082 | 0.90 | 0.65 |
| | $r$ | | 0.1 | 0.3 | 0.2 | | | | |
| 4 | $\pi$ | 1.0 | | | | 1.0 | 0.0189 | 0.60 | 0.95 |
| | $r$ | 0.0 | | | | 0.4 | | | |
| 5 | $\pi$ | | | | 1.0 | 0.9 | −0.1832 | 0.90 | 0.95 |
| | $r$ | | | | 0.2 | 0.3 | | | |
| 6 | $\pi$ | | 0.8 | 1.0 | | | 0.9343 | 0.70 | 0.95 |
| | $r$ | | 0.1 | 0.2 | | | | | |
| 7 | $\pi$ | 0.9 | | | | | −0.3339 | 0.40 | 0.85 |
| | $r$ | 0.3 | | | | | | | |
| 8 | $\pi$ | 0.8 | | | 0.9 | | −0.1006 | 0.90 | 0.95 |
| | $r$ | 0.4 | | | 0.0 | | | | |
| 9 | $\pi$ | | | 1.0 | | 1.0 | 1.0964 | 0.90 | 0.95 |
| | $r$ | | | 0.4 | | 0.3 | | | |
| 10 | $\pi$ | | 0.8 | | | | 0.2996 | 0.60 | 0.95 |
| | $r$ | | 0.4 | | | | | | |

Table 2. True latent ability distribution parameters

| IR | attribute state | ideal response | true $p_l$ | true $\mu_l$ |
|---|---|---|---|---|
| 7 | 00111 | 01001 00010 | 0.09 | −0.057 |
| 9 | 01101 | 00000 10011 | 0.11 | −0.760 |
| 2 | 01111 | 01101 10011 | 0.08 | 0.433 |
| 11 | 10011 | 00011 01100 | 0.11 | −0.392 |
| 3 | 10111 | 01011 01110 | 0.09 | 0.551 |
| 24 | 11000 | 10000 01001 | 0.11 | −0.727 |
| 4 | 11011 | 10011 01101 | 0.10 | −0.964 |
| 5 | 11101 | 10010 11011 | 0.10 | 0.584 |
| 6 | 11110 | 10100 11101 | 0.11 | 0.448 |
| 1 | 11111 | 11111 11111 | 0.10 | 0.762 |
| $\sigma$ | | | | 1.00 |

Table 3. Estimated item parameters

| item | | attr. 1 | attr. 2 | attr. 3 | attr. 4 | attr. 5 | $b_i$ | $c_i$ | $d_i$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | $\pi$ | 1.00 | 0.72 | | | | −0.533 | 0.80 | 0.97 |
| | $r$ | 0.17 | 0.37 | | | | | | |
| 2 | $\pi$ | | | 1.00 | 1.00 | 0.74 | 0.219 | 0.90 | 0.93 |
| | $r$ | | | 0.43 | 0.28 | 0.05 | | | |
| 3 | $\pi$ | | 1.00 | 1.00 | 0.69 | | −0.009 | 0.70 | 0.65 |
| | $r$ | | 0.05 | 0.49 | 0.07 | | | | |
| 4 | $\pi$ | 1.00 | | | | 1.00 | 0.102 | 0.60 | 0.94 |
| | $r$ | 0.00 | | | | 0.38 | | | |
| 5 | $\pi$ | | | | 1.00 | 0.90 | −0.272 | 0.90 | 0.97 |
| | $r$ | | | | 0.22 | 0.26 | | | |
| 6 | $\pi$ | | 1.00 | 0.83 | | | 1.165 | 0.80 | 0.85 |
| | $r$ | | 0.07 | 0.20 | | | | | |
| 7 | $\pi$ | 0.86 | | | | | −0.477 | 0.50 | 0.66 |
| | $r$ | 0.17 | | | | | | | |
| 8 | $\pi$ | 1.00 | | | 0.70 | | −0.390 | 0.90 | 0.98 |
| | $r$ | 0.50 | | | 0.00 | | | | |
| 9 | $\pi$ | | | 1.00 | | 0.97 | −0.090 | 0.30 | 0.95 |
| | $r$ | | | 0.33 | | 0.27 | | | |
| 10 | $\pi$ | | 0.8 | | | | 0.770 | 0.80 | 0.85 |
| | $r$ | | 0.39 | | | | | | |

Table 4. True and estimated ability distribution parameters

| IR | attribute state | ideal response | true $p_l$ | est. $p_l$ | true $\mu_l$ | est. $\mu_l$ | |
|---|---|---|---|---|---|---|---|
| 7 | 00111 | 01001 00010 | 0.09 | 0.0631 | −0.057 | 0.1933 | |
| 9 | 01101 | 00000 10011 | 0.11 | 0.0975 | −0.760 | −0.6773 | |
| 2 | 01111 | 01101 10011 | 0.08 | 0.1037 | 0.433 | −0.0190 | |
| 11 | 10011 | 00011 01100 | 0.11 | 0.1170 | −0.392 | −0.1286 | |
| 3 | 10111 | 01011 01110 | 0.09 | 0.1033 | 0.551 | 0.5177 | |
| 24 | 11000 | 10000 01001 | 0.11 | 0.1175 | −0.727 | −0.5631 | |
| 4 | 11011 | 10011 01101 | 0.10 | 0.0930 | −0.964 | −0.7296 | |
| 5 | 11101 | 10010 11011 | 0.10 | 0.0837 | 0.584 | 0.5579 | |
| 6 | 11110 | 10100 11101 | 0.11 | 0.1004 | 0.448 | 0.9197 | |
| 1 | 11111 | 11111 11111 | 0.10 | 0.0940 | 0.762 | 0.8984 | |
| est. $\sigma$ | | | | | | | 0.931 |

From Table 4, we notice that the true and estimated mixing probabilities are close, while the estimated $\mu_l$ values are often not close to the true $\mu_l$. Because of the way the $\mu_l$'s function in the likelihood through the latent ability distribution, it is relatively more difficult to estimate them accurately, especially when the possibility of relatively flat likelihood surface exists. Taking this into account, we think the estimated $\mu_l$ values are satisfactory (see also the comment below on the comparison between the likelihood at the true and estimated parameters). Note that we start the EM algorithm run with equal mixing probabilities $\left(\frac{1}{24}\right)$ as initial values for all the 24 possible attribute classes. From Table 4, we see that our procedure selects the right 10 classes, the estimated mixing probabilities for the other 14 are all approximately 0 (with an average of 0.0019), as desired. Hence their estimated values are not given.

Because of Proposition 2, different sets of $\pi$'s and $r$'s can generate the same set of $S_{\underline{\alpha},i}$'s. In our estimation procedure, remember that for an item $i$ that requires $K_i$ attributes we (arbitrarily) fix the first $K_i - 1$ of the $\pi$'s at 1 to reduce the indeterminacy among the $\pi$'s and $r$'s. As a result, the parameter estimates for some items may appear to be far away from their true values. Because of this problem, to determine the estimation accuracy of item parameters, we need to instead compare the estimated $S_{\underline{\alpha},i}$'s with their true values using the estimated and true values of $\pi$'s and $r$'s. From Table 3 it is clear that for items 4, 5, 7, 9, and 10 the estimates of $\pi$'s and $r$'s are close to their true values. Consequently for these items the estimated $S_{\underline{\alpha},i}$'s will be close to the true values. Because the estimates of $\pi$'s and $r$'s are far off the true values for items 1, 2, 3, 6, and 8, Tables 5 and 6 show the comparisons of the estimated $S_{\underline{\alpha},i}$'s and their true values for these items. Table 5 compares the true values for $S_{\underline{\alpha},i}$'s with their estimates for items 1, 6, and 8, while Table 6 compares the true $S_{\underline{\alpha},i}$ with their estimates for items 2 and 3. From Tables 5 and 6, we see that the estimated values of $S_{\underline{\alpha},i}$'s are close to their true values for all the possible attribute states for these items (the average absolute deviation between the true and estimated $S_{\underline{\alpha},i}$ is 0.0186 for Table 5 and 0.0295 for Table 6), even though the individual $\pi$ and $r$ estimates are not close to the true values. Note that in Tables 5 and 6, when denoting the attribute states, we list only the attributes required by the item. For example, in Table 5 attribute state 10 is really $(1, 0, \star, \star, \star)^T$ for item 1, while it is $(\star, 1, 0, \star, \star)^T$ for item 6, where $\star$ denotes it can be either 1 or 0.

BEST COPY AVAILABLE

Table 5. True and estimated $S_{\underline{\alpha},i}$'s for items 1, 6, 8

| | item 1 | | item 6 | | item 8 | |
|---|---|---|---|---|---|---|
| attribute state | true $S_{\underline{\alpha},i}$ | estimated $S_{\underline{\alpha},i}$ | true $S_{\underline{\alpha},i}$ | estimated $S_{\underline{\alpha},i}$ | true $S_{\underline{\alpha},i}$ | estimated $S_{\underline{\alpha},i}$ |
| 11 | 0.720 | 0.7200 | 0.800 | 0.8300 | 0.720 | 0.7000 |
| 10 | 0.360 | 0.3700 | 0.160 | 0.2000 | 0.000 | 0.0000 |
| 01 | 0.080 | 0.1224 | 0.100 | 0.0581 | 0.360 | 0.3500 |
| 00 | 0.040 | 0.0629 | 0.020 | 0.0140 | 0.000 | 0.0000 |

Table 6. True and estimated $S_{\underline{\alpha},i}$'s for items 2 and 3

| item 2 | | | item 3 | | |
|---|---|---|---|---|---|
| attribute state | true $S_{\underline{\alpha},i}$ | estimated $S_{\underline{\alpha},i}$ | attribute pattern | true $S_{\underline{\alpha},i}$ | estimated $S_{\underline{\alpha},i}$ |
| 111 | 0.720 | 0.7400 | 111 | 0.640 | 0.6900 |
| 110 | 0.080 | 0.0500 | 110 | 0.160 | 0.0700 |
| 101 | 0.216 | 0.2072 | 101 | 0.240 | 0.3381 |
| 011 | 0.360 | 0.3182 | 011 | 0.064 | 0.0345 |
| 100 | 0.024 | 0.0140 | 100 | 0.060 | 0.0343 |
| 010 | 0.040 | 0.0215 | 010 | 0.016 | 0.0035 |
| 001 | 0.108 | 0.0891 | 001 | 0.024 | 0.0169 |
| 000 | 0.012 | 0.0060 | 000 | 0.006 | 0.0017 |

By examining Tables 1 and 3, we see that the parameter estimates of $b_i$, $c_i$, and $d_i$ are not close to their true values. But remember when $d_i$ is close to 1, we cannot expect to accurately estimate $b_i$ and $c_i$ separately because the data contains little information about these parameters (i.e., the likelihood surface is rather flat). That is, we are close to a condition of unidentifiability of the $b_i$ and $c_i$. For those items with $d_i$ close to 1, as discussed earlier a comparison between the true and estimated $b_i - 2c_i$ is appropriate, and Table 7 shows they are quite close.

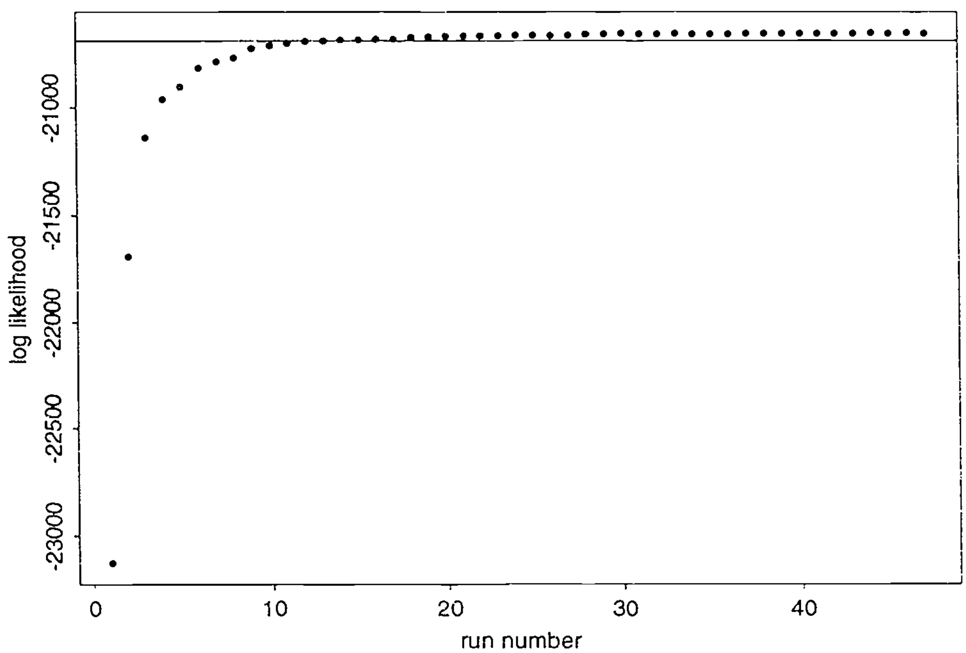
Table 7. Comparison of true and estimated item parameters

| item | true $b_i$ | est. $b_i$ | true $c_i$ | est. $c_i$ | true $d_i$ | est. $d_i$ | true $b_i - 2c_i$ | est. $b_i - 2c_i$ |
|---|---|---|---|---|---|---|---|---|
| 1 | −0.3497 | −0.533 | 0.90 | 0.80 | 0.90 | 0.97 | −2.1497 | −2.133 |
| 2 | −0.1929 | 0.219 | 0.70 | 0.90 | 0.95 | 0.93 | −1.5929 | −1.581 |
| 4 | 0.0189 | 0.102 | 0.60 | 0.60 | 0.95 | 0.94 | −1.1811 | −1.098 |
| 5 | −0.1832 | −0.272 | 0.90 | 0.90 | 0.95 | 0.97 | −1.9832 | −2.072 |
| 6 | 0.9343 | 1.165 | 0.70 | 0.80 | 0.95 | 0.85 | −0.4657 | −0.435 |
| 8 | −0.1006 | −0.390 | 0.90 | 0.90 | 0.95 | 0.98 | −1.9006 | −2.190 |
| 9 | 1.0964 | −0.090 | 0.90 | 0.30 | 0.95 | 0.95 | −0.7036 | −0.690 |
| 10 | 0.2996 | 0.770 | 0.60 | 0.80 | 0.95 | 0.85 | −0.9004 | −0.830 |

Remember the ultimate goal of our estimating the model parameters is to enable us to classify the examinees cognitively. Since we are using the marginal maximum likelihood approach to estimate the model parameters (the calibration step, preliminary to the classification step), another way (the right way) to look at the estimation accuracy of our model calibration procedure is to compare the log likelihood values given at the true and estimated parameters. Because of the possibility of a relatively flat likelihood surface in certain locations of the model parameter space, different parameter sets might give approximately the same likelihood. However, if we use the estimated likelihood as input to an examinee cognitive classification procedure, it is the likelihood value rather than the estimated parameter values that is central; thus non-influential differences in estimated parameter values are irrelevant.

For the model we are considering, Figure 1 gives the plot of values of the log likelihood from each of the EM cycles of a particular run using our estimation program. The horizontal line in the figure corresponds to the true log likelihood, which is −20688.66. From Figure 1, we can see the log likelihood values for the first several EM cycles are rapidly approaching the true log likelihood. After 13 or so cycles the log likelihood value is already quite stable. For the last 15 or so cycles the log likelihood values are increasing very slowly. The estimated log likelihood from the final EM cycle is -20629.32, larger than but close to the true log likelihood (the reason why the estimated log likelihood is larger than the true log likelihood for this particular run might be due to the combination effect of estimation error and the randomness we have introduced in generating the data).

Figure 2. Plot of the Log Likelihood Values from the EM Cycles




# 5 Summary

The important need for test analysis methods that extract cognitive information useful to the practitioners from ordinary tests is widely recognized and is a topic of vigorous research in psychometrics and cognitive psychology. Such methods and underlying theory should be applicable to tests that are in common use today, as well as in the future to specially constructed diagnostic instruments based upon cognitive theory, in many cases computer administered. The goal of developing the Unified Model was to be able to determine, on the basis of a simple test, what the cognitive strengths and weaknesses of an examinee are, relative to a list of cognitive attributes of interest in the particular educational setting of the test.

The Unified Model is theoretically appealing relative to other cognitive diagnostic models, but because of its structural complexity, there is not yet estimation package available. In this paper, we have proposed an estimation procedure for the Unified Model and have shown that it is not only computationally feasible but effective. With an effective estimation procedure for the Unified Model, we can calibrate the model, and thus classify and estimate examinee latent abilities, thereby extracting useful cognitive information about the test as well as the examinees.

## References


**Bartholomew, D.** (1987). *Latent variable models and factor analysis*. London: Charles Griffin.

**Bock, R. D., and Aitkin, M.** (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46, 443-59.

**Dempster, A. P., Laird, N. M., and Rubin, D. B.** (1977). Maximum Likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

**DiBello, L. V., Stout, W. F., and Roussos, L. A.** (1993). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, and R. L. Brennan (Eds.), *Cognitively diagnostic assessment*. Hillsdale, NJ: Lawrence Erlbaum.

**Haertel, E.** (1990). Continuous and discrete latent structure models for item response data. *Psychometrika*, 55, 477-94.

**Jiang, H.** (1996). *Theory and Applications of Computational Statistics in Cognitive Diagnosis and IRT Modeling*. Ph.D. Thesis, University of Illinois at Urbana-Champaign.

**Lazersfeld, P. F., and Henry, N.W.** (1968). *Latent structure analysis*. New York: Houghton-Mifflin.

**Michalewicz, Z.** (1994). *Genetic algorithms + data structures = evolution programs*. Berlin: Springer-Verlag.

**McCormick, G.** (1967). Second order conditions for constrained minima. *SIAM Journal on Applied Mathematics*, 15, 641-52.

**Snow, R., and Lohman, D.** (1993). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational measurement*. New York: American Council on Education, Macmillan.

**Takane, Y., and de Leeuw, J.** (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 52, 393-408.

**Tatsuoka, K. K.** (1984). Caution indices based on item response theory. *Psychometrika*, 49, 95-110.

**Tatsuoka, K. K.** (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, 12, 55-73.

**Tatsuoka, K. K.** (1990). Toward an integration of item-response theory and cognitive error diagnoses. In N. Frederiksen, R. L. Glaser, A. M. Lesgold, and M. G. Shafto (Eds.), *Diagnostic monitoring of skill and knowledge acquisition.* Hillsdale, NJ: Lawrence Erlbaum.

**Tatsuoka, K. K., and Tatsuoka, M. M.** (1987). Bug distribution and pattern classification. *Psychometrika*, 52, 193-206.

**Yamamoto, K.** (1987). *A model that combines IRT and latent class models.* Ph.D. Thesis. University of Illinois at Urbana-Champaign.